# ORIGINAL ARTICLE

# Genome-wide haplotype association study identifies the *FRMD4A* gene as a risk locus for Alzheimer's disease

J-C Lambert[1,2,3,47], B Grenier-Boley[1,2,3,47], D Harold[4], D Zelenika[5], V Chouraki[1,2,3], Y Kamatani[5,6], K Sleegers[7,8], MA Ikram[9], M Hiltunen[10], C Reitz[11], I Mateo[12], T Feulner[13], M Bullido[14], D Galimberti[15], L Concari[16], V Alvarez[17], R Sims[4], A Gerrish[4], J Chapman[4], C Deniz-Naranjo[18], V Solfrizzi[19], S Sorbi[20], B Arosio[21], G Spalletta[22], G Siciliano[23], J Epelbaum[24], D Hannequin[25], J-F Dartigues[26], C Tzourio[27,28], C Berr[29], EMC Schrijvers[30], R Rogers[11], G Tosto[11], F Pasquier[3,31], K Bettens[7,8], C Van Cauwenberghe[7,8], L Fratiglioni[32,33], C Graff[33,34], M Delepine[5], R Ferri[35], CA Reynolds[36], L Lannfelt[37], M Ingelsson[37], JA Prince[38], C Chillotti[39], A Pilotto[40], D Seripa[40], A Boland[5], M Mancuso[23], P Bossù[22], G Annoni[21], B Nacmias[20], P Bosco[35], F Panza[19], F Sanchez-Garcia[18], M Del Zompo[41], E Coto[17], M Owen[4], M O'Donovan[4], F Valdivieso[14], P Caffara[16], E Scarpini[15], O Combarros[12], L Buée[2,42], D Campion[25], H Soininen[10], M Breteler[43,44], M Riemenschneider[13], C Van Broeckhoven[7,8], A Alpérovitch[27,28], M Lathrop[5,6], D-A Trégouët[45,46], J Williams[4] and P Amouyel[1,2,3,31], EADI consortium[48], GERAD consortium[48]

[1]INSERM, U744, Lille, France; [2]Institut Pasteur de Lille, Lille, France; [3]Université Lille-Nord de France, Lille, France; [4]Medical Research Council (MRC) Centre for Neuropsychiatric Genetics and Genomics, Neurosciences and Mental Health Research Institute, Department of Psychological Medicine and Neurology, School of Medicine, Cardiff University, Cardiff, UK; [5]Centre National de Genotypage, Institut Genomique, Commissariat à l'énergie Atomique, Evry, France; [6]Fondation Jean Dausset-CEPH, Paris, France; [7]Neurodegenerative Brain Diseases group, Department of Molecular Genetics, VIB, Antwerpen, Belgium; [8]Laboratory of Neurogenetics, Institute Born-Bunge, University of Antwerp, Antwerpen, Belgium; [9]Departement of Epidemiology and Radiology, Erasmus MC, Rotterdam; [10]Department of neurology, University of Eastern Finland and Kuopio University Hospital, Kuopio, Finland; [11]G.H. Sergievsky Center, Columbia University, New York, NY, USA; [12]Service of Neurology, University Hospital Marques de Valdecilla, University of Cantabria, Santander, Spain; [13]Department of Psychiatry and Psychotherapy, Universitätsklinikum des Saarlandes, Universität des Saarlandes Saarbruecken, Germany; [14]Centro de Biologia Molecular Severo Ochoa (UAM-CSIC) and CIBERNED, Universidad Autonoma, Madrid, Spain; [15]Department of Neurological Sciences, Dino Ferrari Center, University of Milan, Fondazione Cà Granda, IRCCS Ospedale Maggiore Policlinico, Milan, Italy; [16]Department of Neuroscience, University of Parma, Parma, Italy; [17]Genética Molecular-Laboratorio de Medicina, Hospital Universitario Central Asturias, Oviedo, Spain; [18]Servicio de Inmunología. Hospital Unviersitario de Gran Canaria Dr Negrín. Bco, Las Palmas de Gran Canaria, Spain; [19]Department of Geriatrics, Center for aging brain, Memory unit, University of Bari, Bari, Italy; [20]Department of neurological and psychiatric Sciences, University of Florence, Florence, Italy; [21]Department of Internal Medicine, Geriatric Unit, Fondazione IRCCS Cà Granda Ospedale Maggiore Policlinico, Università degli Studi di Milano, Milan, Italy; [22]Clinical and Behavioral Neurology, IRCCS Fondazione Santa Lucia, Roma, Italy; [23]Neurological clinic, University of Pisa, Pisa, Italy; [24]INSERM UMR 894, Paris Descartes University, Paris, France; [25]Inserm U614, Faculté de Médecine-Pharmacie de Rouen, Rouen, France; [26]Inserm U897, Victor Segalen University, Bordeaux, France; [27]Inserm U708, Paris, France; [28]UPMC University Paris 06, Paris, France; [29]Inserm U1061, Hôpital La Colombière, Montpellier, France; [30]Department of Epidemiology and Neurology, Erasmus MC, Rotterdam; [31]CHRU de Lille, Lille, France; [32]Aging Reasearch Center, Department NVS, Karolinska Institutet and Stockholm University, Stockholm, Sweden; [33]Department of Geriatric Medicine, Karolinska University Hospital, Stockholm, Sweden; [34]KI-Alzheimer's Disease Reseach Center, Department NVS, Karolinska Institutet, KIADRC, Stockholm, Sweden; [35]IRCCS Associazione Oasi Maria SS, Institute for Research on Mental Retardation and Brain Aging, Troina (EN), Italy; [36]Department of Psychology, University of California at Riverside, Riverside, CA, USA; [37]Department of Public health and Caring Sciences, Rudbeck Laboratory, Uppsala University, Uppsala, Sweden; [38]Department of Medical Epidemiology and Biostatistics, Karolinska Institute, Stockholm, Sweden; [39]Unit of Clinical Pharmacology, Teaching Hospital of Cagliari, Cagliari, Italy; [40]Geriatric Unit & Gerontology-Geriatrics Research Laboratory, Department of Medical Sciences, I.R.C.C.S. 'Casa Sollievo della Sofferenza', San Giovanni Rotondo, Italy; [41]Section of Clinical Pharmacology, Department of Neurosciences 'B.B. Brodie', University of Cagliari, Cagliari, Italy; [42]Inserm U837, Lille, France; [43]Department of Epidemiology, Erasmus MC, Rotterdam; [44]DZNE, Bonn, Germany; [45]INSERM UMR_S 937, Paris, France and [46]ICAN Institute of Cardiometabolism and Nutrition, Université Pierre & Marie Curie, Paris, France

Correspondence: Dr J-C Lambert, INSERM U744, Institut Pasteur de Lille, BP 245, 1 rue du professeur Calmette, 59019 Lille cedex, France.
E-mail: jean-charles.lambert@pasteur-lille.fr
[47]These authors contributed equally to this work.
[48]A full list of members is provided in the Supplementary Note.

Recently, several genome-wide association studies (GWASs) have led to the discovery of nine new loci of genetic susceptibility in Alzheimer's disease (AD). However, the landscape of the AD genetic susceptibility is far away to be complete and in addition to single-SNP (single-nucleotide polymorphism) analyses as performed in conventional GWAS, complementary strategies need to be applied to overcome limitations inherent to this type of approaches. We performed a genome-wide haplotype association (GWHA) study in the EADI1 study (*n* = 2025 AD cases and 5328 controls) by applying a sliding-windows approach. After exclusion of loci already known to be involved in AD (*APOE*, *BIN1* and *CR1*), 91 regions with suggestive haplotype effects were identified. In a second step, we attempted to replicate the best suggestive haplotype associations in the GERAD1 consortium (2820 AD cases and 6356 controls) and observed that 9 of them showed nominal association. In a third step, we tested relevant haplotype associations in a combined analysis of five additional case–control studies (5093 AD cases and 4061 controls). We consistently replicated the association of a haplotype within *FRMD4A* on Chr.10p13 in all the data set analyzed (OR: 1.68; 95% CI: (1.43–1.96); $P = 1.1 \times 10^{-10}$). We finally searched for association between SNPs within the *FRMD4A* locus and A$\beta$ plasma concentrations in three independent non-demented populations (*n* = 2579). We reported that polymorphisms were associated with plasma A$\beta$42/A$\beta$40 ratio (best signal, $P = 5.4 \times 10^{-7}$). In conclusion, combining both GWHA study and a conservative three-stage replication approach, we characterised *FRMD4A* as a new genetic risk factor of AD.

## Introduction

The identification of genes involved in monogenic forms of Alzheimer's disease (AD) has significantly contributed to our knowledge of the disease mechanisms. The causal links between mutations, the functions of the mutated genes (*APP*, *PS1* and *PS2*) and disease development prompted a pathophysiological hypothesis, which radically changed our understanding of AD: the amyloid cascade hypothesis.[1] The systematic association of pathogenic mutations with changes in APP metabolism and, more particularly, a relative overproduction of A$\beta$42 peptides indicates that this metabolism is at the heart of the disease process (at least in the monogenic forms of the disease). The overproduction of these neurotoxic peptides is supposed to lead to or accentuate neuron-to-neuron propagation of the $\tau$ pathology (leading to neuronal death) by an unknown mechanism.[2]

By analogy, it was expected that the characterization of genetic factors involved in the common forms of AD (that is, lacking classical Mendelian inheritance), the most frequent form of the disease, should also help to better understand the AD physiopathological process. However, the characterization of these genetic factors has encountered significant difficulties. Until 2009, the *APOE* (apolipoprotein E) gene was the only globally valid genetic determinant of AD to have been unambiguously identified in 15 years of intensive research.[3,4]

As with other multifactorial diseases, this systematic inability to detect new genetic determinants has prompted more comprehensive investigations using genome-wide association studies (GWASs). We and others performed five large GWASs in this field and reported that the *CLU* (clusterin), *PICALM* (phosphatidylinositol-binding clathrin assembly protein), *CR1* (complement component (3b/4b) receptor 1), *BIN1* (bridging integrator 1), *ABCA7* (ATP-binding cassette, sub-family A, member 7), *MS4A* (membrane spanning 4A) cluster, *EPHA1* (ephrin type-A receptor 1), *CD33* (differentiation antigen 33) and *CD2AP* (CD2-associated protein) genes were associated with the AD risk.[5–9] Most of these susceptibility genes have been already systematically replicated in Caucasians in large case–control studies and in families.[10]

However, our understanding of the AD genetics is far away to be complete and strong efforts have still to be done. At this level, classical GWAS approaches present an important limitation with systematic application of a conventional, highly conservative Bonferroni correction leading to select only the most statistically significant associations (commonly, $P < 1 \times 10^{-8}$). This involves the risk of rejecting biologically valid hypotheses on purely statistical grounds, that is, false negatives.

To partly handle some of this limitation, other complementary approach consists in extracting pertinent information from single-nucleotide polymorphisms (SNPs) nominally associated with the risk of developing AD in GWAS by using complex statistical and bioinformatics multiple-SNP analyses such as genome-wide haplotype association (GWHA) study.[11] We adopted this GWHA strategy to the AD genetic susceptibility through a three-step approach.

## Materials and methods

### Population description

The main characteristics of the populations used for the GWHA study are described in the Supplementary

Notes and Supplementary Table 1. All AD cases met the criteria for either probable AD (NINCDS-ADRDA, DSM-IV)[12] or definite AD (CERAD).[13] All elderly controls were screened for dementia using the MMSE or the AD Assessment Scale-cognitive subscale and were determined to be free from dementia at neuropathological examination or had a Braak score of 2.5 or below. All subjects or, in those with substantial cognitive impairment, a caregiver, legal guardian, or other proxy gave written informed consent for participation in this study. The study protocols for all populations were reviewed and approved by the appropriate Institutional review boards of each country.

The main characteristics of the populations (non-demented individuals) used for the Aβ plasma study (3C study, Rotterdam and CHS) are described in the Supplementary Notes and Supplementary Table 2.[14–16]

*Genotyping*
Participants in the French GWA study (including the 3C study) were genotyped using an Illumina 610–quad array (Illumina, San Diego, CA, USA). Quality control and analytical parameters have been described in detail elsewhere.[5] The GERAD participants were genotyped using an Illumina 610–quad array (Illumina), a HumanHap550 array or a HumanHap300 array (Illumina). Again, QC and analytical parameters have been described in detail elsewhere.[6] The Rotterdam and CHS studies were genotyped using the Affymetrix 500K array (Affymetrix, Santa Clara, CA, USA). Again, QC and analytical parameters have been described in detail elsewhere.[7]

In the European populations (stage 3), genotyping was performed using Sequenom assays at the exception of the German population genotyped with a 610K ILLUMINA chip (Illumina). The primer and probe sequences used in the genotyping assays are available upon request. In order to avoid any genotyping bias, cases and controls were randomly mixed while genotyping and laboratory personnel were blinded to case/control status. The genotyping success rate was at least 95%. Departure from Hardy–Weinberg equilibrium was observed for rs2446581 in the Swedish control and case samples ($P = 9.1 \times 10^{-7}$ for the whole population). The Swedish sample was consequently excluded from further analysis, as haplotype analyses that do not comply with Hardy–Weinberg equilibrium are likely to bias observations.

*Statistical analyses*

*Missing age or gender data.* Any individuals with missing age or gender data were excluded. This gave a maximum of 2025 AD cases and 5328 controls in step 1, 2820 AD cases and 6356 controls in step 2 and 5093 AD cases and 4061 controls in step 3.

*Detection of haplotype effects using a sliding-window approach.* This approach has been fully described elsewhere.[11,17] Briefly, the search for haplotype effects

was carried out by applying a sliding-window approach[18,19] to the French GWA data set for each chromosome. After excluding SNPs not in Hardy–Weinberg equilibrium or with a minor allele frequency (MAF) < 0.02, the first step of the strategy was to eliminate part of the redundancy between SNPs by using haplotype-tagging SNPs (htSNPs). For this, the same binning procedure as in[11] was used: within each bin of 10 adjacent SNPs, we identified a minimal set of htSNPs that were able to characterize more than 95% of the inferred haplotypes with estimated frequency greater than 0.02. Once a bin had been characterized by a set of htSNPs, the same strategy was applied to the bin composed of the next 10 adjacent SNPs. The final set of htSNPs ($n = 287\,956$) was then fed into the sliding-windows approach.

Given a window of 10 htSNPs, the search for the most informative and parsimonious haplotype configuration in terms of disease prediction was performed for all possible 1 to 4 loci combinations of not necessarily adjacent SNPs. We used a strategy based on Akaike's information criterion, which has been previously described for candidate gene haplotype analysis.[19,20] It relies on the stochastic expectation–maximisation module[21] in THESIAS software.[22] If required, missing genotypes were inferred by applying,[23] multiple imputation.[21] In all, 37 330 050 combinations were investigated in our genome scan and this investigation was conducted thanks to the use of the grid technology developed by the European Grid Infrastructure (http://www.egi.eu).[20] This technology enables several thousand computations to be run in parallel on a large number of different CPUs. The sliding-window haplotype approach was developed into a GridHaplo grid package for the EGI grid (http://genecanvas.ecgene.net).[11]

*Replication of haplotype effects.* Regions with 'window P-values' (i) below $10^{-5}$ and (ii) 100 times smaller than the smallest single-locus P-value (including all SNPs and not only htSNPs) were analyzed in terms of replication in the GERAD1 data set by using THESIAS software, with systematic adjustment for age and gender.

*Association of FRMD4A SNPs with Aβ plasma concentrations.* In each center, Aβ plasma variables are normally distributed. We excluded prior analyses all samples with a +/− two s.d. values in order to avoid potential associations driven by extreme observations. Finally, each quantitative variable was transformed into a z-score (equal to (observed value minus the sample mean), divided by the sample s.d.). The association between the $A\beta_{1-40}$, $A\beta_{1-42}$ and $A\beta_{1-42}/A\beta_{1-40}$ z-scores on one hand and imputed *FRMD4A* imputed SNPs on the other (see below) were assessed using a general linear model under an additive model adjusted for age, center and gender.

We used inverse-variance weighting (also known as fixed-effects meta-analysis) to investigate the homogeneity of haplotype effects from one study

to another and to provide meta-analysed, age- and gender-adjusted ORs for haplotype effect estimates in the seven studies. A similar strategy was used to provide meta-analysed, age- and gender-adjusted association levels between Aβ plasma z-scores and *FRMD4A* SNPs.

*Imputation analyses*
We imputed SNPs by using MaCH (http://www.sph.umich.edu/csg/abecasis/mach/index.html) and mini-mac software (http://genome.sph.umich.edu/wiki/minimac). The reference haplotype data are provided by the MaCH website, which was built for the combined Caucasian populations as part of the 1000 genomes project. In our data set, all individuals were genotyped on the same platform (the Illumina Human660W-Quad Beadchip) and we used 492 941 observed SNP genotypes that passed quality filters as follows: genotyping call rate $\geqslant 98\%$, Hardy–Weinberg Equilibrium Test $P$ value $\geqslant 1 \times 10^{-6}$ and MAF $\geqslant 1\%$. We first inferred haplotype combinations of each individual using the 'phase' option in the MaCH program and then imputed them with minimac. As minimac is a newly developed software tool, we compared the correlation between the imputed genetic dosage from minimac and those from the standard MaCH program for SNPs in chromosome 22. The results were very similar (data not shown). Doses for 7 704 555 million SNPs with a MAF $> 0.01$ were available from the French GWA data set using the 1000 genomes data set. We selected 2538 SNPs within the Chromosome 10p13 locus of interest (chromosome10: 13 655 705–14 402 866) and evaluated their associations with AD risk in an additive logistic regression model adjusted for age, gender and disease status. A graphic representation was then generated with Locus-zoom software (http://csg.sph.umich.edu/locuszoom/).

In the 3C, Rotterdam and CHS cohorts, we used the genotype data to impute to the 2.5 million non-monomorphic, autosomal SNPs described in HapMap II (CEU population) as described elsewhere.[7] We selected 1486 SNPs (MAF $> 0.01$) within the Chromosome 10p13 locus of interest (chr10:13 655 705– 14 402 866) and evaluated their associations with Aβ plasma concentrations as described above. Again, a graphic representation was then generated with Locus-zoom software (http://csg.sph.umich.edu/locuszoom/).

## Results

We developed a three-step approach. In the first step, the French GWA study (EADI1 for European Alzheimer's initiative 1),[5] including 2025 AD cases and 5328 controls was used to select regions with potential haplotype associations with AD. Following a sliding-windows approach (see Material and methods), we applied two *a priori* criteria to select loci of interest as previously described:[11] level of association with a $P$-value (i) below $10^{-5}$ and (ii) at least 100 times smaller than the smallest single-SNP $P$ value observed in the corresponding locus. We were able to detect loci already known to be involved in AD (the *APOE*, *BIN1* and *CR1* locus)[5–9] from previous GWASs. The obtained signal was systematically highly stronger than the one observed for each SNP taken separately (Table 1). After exclusion of these loci, we retained 91 regions of interest.

In the second step, we replicated these haplotype associations in the GWA database from another consortium involved in the study of AD genetic susceptibility, the GERAD1 consortium,[6] including 2820 AD cases and 6356 controls. All 91 regions were available for investigation and 9 of them showed nominal association ($P < 0.05$; Table 2). We decided to further investigate two regions showing the same best haplotypes associated with AD risk in both EADI1 and GERAD1, with similar magnitude and direction of association (Table 2). These two loci, not previously detected in single-SNP GWAS analyses, were located on chromosomes 6p21 and 10p13.

**Table 1** Best haplotype combination observed in *CR1*, *BIN1* and *APOE* and comparison with the association obtained for single-SNP analysis within these loci

| Locus | hTag-SNPs | Global haplotype P-value | Best associated hTag-SNP | | Best associated SNP within the locus | |
|---|---|---|---|---|---|---|
| | | | RS number | Single P-value for hTag-SBP | RS number | P value_single (CNG) |
| CR1 | rs3818361 + rs1323720 + rs7527798 + rs2761424 | 4.12E−10 | rs3818361 | 2.6E−06 | rs3818361 | 2.6E−06 |
| BIN1 | rs17014873 + rs749008 + rs13031703 + rs744373 | 3.13E−09 | rs744373 | 1.04E−03 | rs10194375 | 2.4E−04 |
| APOE | rs157580 + rs8106922 + rs405509 + rs439401 | 9.51E−184 | rs157580 | 5.4E−32 | rs2075650 | 3.32E−126 |

Abbreviations: APOE, apolipoprotein E; BIN1, bridging integrator 1; CR1, complement component (3b/4b) receptor 1; hTag-SNPs, haplotype-tagging SNPs; SNP, single-nucleotide polymorphism.

**Table 2** List of the haplotype combinations identified in the French GWA study and replicated at $P<0.05$ in GERAD

| Chromosome | SNPs used in haplotypes | First position | Final position | EADI1 | | | | | GERAD1 | | | | | Best meta-analysed haplotype association | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Best haplotype | OR | 95% CI | Best haplotype P value | Gloabl P value | Best haplotype | OR | 95% CI | Best haplotype P value | Gloabl P value | Best haplotype | OR | 95% CI | Best haplotype P value |
| 2 | rs6715901 + rs2366913 + rs10497527 + rs1489479 | 179 359 199 | 179 416 922 | ACGT | 0.69 | (0.58–0.83) | 4.30E−05 | 2.77E−06 | GTGT | 0.56 | (0.37–0.87) | 9.42E−03 | 4.70E−02 | ACGT | 0.80 | (0.69–0.91) | 1.23E−03 |
| 3 | rs696236 + rs7641821 + rs3911778 + rs1471377 | 64 184 258 | 64 228 008 | GCCG | 1.33 | (1.16–1.52) | 2.85E−05 | 4.65E−06 | GTCA | 0.68 | (0.53–0.87) | 2.34E−03 | 2.43E−02 | GTCA | 0.82 | (0.71–0.95) | 8.15E−03 |
| 3 | rs830637 + rs7612407 | 71 753 155 | 71 769 998 | TA | 0.54 | (0.41–0.71) | 1.13E−05 | 5.27E−06 | TA | 1.50 | (1.06–2.13) | 2.30E−02 | 3.04E−02 | CA | 1.25 | (1.08–1.45) | 3.55E−03 |
| 4 | rs12645972 + rs1466614 + rs6812414 | 37 966 941 | 37 999 202 | TAA | 1.40 | (1.21–1.62) | 4.63E−06 | 9.39E−06 | CAC | 0.82 | (0.71–0.94) | 4.33E−03 | 1.90E−02 | TAA | 1.23 | (1.10–1.38) | 1.97E−04 |
| 5 | rs11740632 + rs7722373 + rs7723349 + rs6893752 | 60 177 002 | 60 410 669 | AAGG | 1.17 | (1.06–1.29) | 2.65E−03 | 8.89E−06 | AATG | 0.81 | (0.71–0.93) | 2.41E−03 | 1.55E−02 | AATG | 0.87 | (0.80–0.95) | 1.25E−03 |
| 6 | rs4711652 + rs2395760 + rs991762 | 40 960 717 | 41 031 388 | GGT | 1.54 | (1.27–1.87) | 1.02E−05 | 7.91E−06 | GGT | 1.51 | (1.16–1.96) | 2.26E−03 | 1.14E−02 | GGT | 1.53 | (1.31–1.79) | 8.08E−08 |
| 6 | rs10946155 + rs1997661 + rs9355578 | 166 644 302 | 166 739 764 | CCT | 1.69 | (1.41–2.02) | 1.26E−08 | 2.42E−06 | TTC | 1.40 | (1.04–1.88) | 2.59E−02 | 4.82E−02 | CCT | 1.45 | (1.26–1.66) | 1.41E−07 |
| 10 | rs7081208 + rs2446581 + rs17314229 | 14 031 871 | 14 056 165 | AAC | 1.80 | (1.40–2.31) | 4.48E−06 | 4.08E−06 | AAC | 1.71 | (1.23–2.37) | 1.36E−03 | 7.39E−03 | AAC | 1.76 | (1.44–2.15) | 2.27E−08 |
| 11 | rs7950171 + rs10834863 + rs276907 + rs2672221 | 3 661 280 | 3 719 744 | TGCT | 0.68 | (0.59–0.77) | 1.21E−08 | 8.75E−07 | TATT | 0.60 | (0.47–0.76) | 4.27E−05 | 5.93E−03 | TGCT | 0.79 | (0.71–0.87) | 2.14E−06 |

Abbreviations: CI, confidence interval; GWA, genome-wide association; OR, odds ratio; SNP, single-nucleotide polymorphism.
Exactly replicated combinations (i.e., the same best haplotypes associated with AD risk in the datasets from the French GWA study and GERAD, with similar magnitudes of association and the same direction of association) are indicated in gray.

The identified AD-associated haplotypes at the 6p21 region were tagged by rs2395760, rs991762 and rs4711652. The rs7081208, rs2446581 and rs17314229 tagged the 10p13 haplotypes. For the Chromosome 6p21 locus, the best association was attributable to the GGT haplotype in both GWASs (OR (Odds Ratio): 1.53; 95% CI: (1.31–1.79); $P = 8.1 \times 10^{-8}$ after adjustment for age and gender when both EADI1 and GERAD1 studies were combined). For the Chromosome 10p13 locus, the highest level of association was attributable to the AAC haplotype (OR: 1.76; 95% CI: (1.44–2.15); $P = 2.3 \times 10^{-8}$ after adjustment for age and gender when both EADI1 and GERAD1 studies were combined. Additional adjustment for the four main principal components did not modify the results, data not shown) (Table 2).

In the third step, the six tagging SNPs were further genotyped in five additional AD case–control studies from Flanders-Belgium (842 cases and 489 controls), Finland (560 cases and 623 controls), Germany (728 cases and 961 controls), Italy (1846 cases and 904 controls) and Spain (1117 cases and 1084 controls) (Supplementary Tables 3 and 4). The same common haplotypes were inferred from the six SNPs in the two loci, with the exception of the AGT haplotype in Chromosome 10p13 in the Italian population (frequency < 0.01 in controls) (Supplementary Table 5 and 6).

In a combined analysis of the five replication data sets, the overall difference in haplotype distribution for the 6p21 locus was not significant between AD cases and controls ($P = 0.13$) and the GGT haplotype was not associated with AD risk (OR: 0.89; 95% CI: (0.74–1.08); $P = 0.23$ after adjustment for age and gender). We therefore considered that the haplotype association in this locus was not confirmed (see Supplementary Table 5).

Conversely, an overall significant difference in haplotype distribution between AD cases and controls

was observed at the 10p13 locus in the replication sample ($P = 4.3 \times 10^{-2}$ after adjustment for age, gender and country) and the AAC haplotype was associated with increased AD risk in the five data sets meta-analysis (OR: 1.55; 95% CI: (1.19–2.00); $P = 9.2 \times 10^{-4}$ after adjustment for age and gender). When the seven data sets (EADI1, GERAD1 and the follow-up studies) were analyzed together, the AAC haplotype had a meta-analysed OR of 1.68 (95% CI: (1.43–1.96); $P = 1.1 \times 10^{-10}$ adjusted for age and gender) when compared with the most frequent GGC haplotype, with no evidence of heterogeneity across the seven countries ($P = 0.92$) (Figure 1). Among the whole sample, nine cases were homozygous for the AAC haplotype versus only five controls (OR: 2.85; 95% CI: (0.88–9.76); $P = 0.09$ with Yates correction), which is consistent with a dose-dependent effect of the AAC haplotype.

Importantly, none of the individual SNPs were associated with AD risk in the total sample with a $P$ value lower than $10^{-5}$. The best meta-analysed SNP was rs2446581 (OR: 1.15; 95% CI: (1.08–1.24); $P = 2.1 \times 10^{-5}$ after adjustment for age and gender; Supplementary Table 7). The hypothesis that the rs2446581 could solely explain the association observed with the 10p13 haplotypes was rejected by use of the likelihood ratio test ($P = 9.1 \times 10^{-7}$).

We also tested whether this haplotype association might be explained by one or more untyped SNPs located nearby the genotyped SNPs. Genotypes for 7 704 555 million SNPs with a MAF $\geqslant 0.01$ were imputed from the French GWA data by using the 1000 genome dataset (http://www.1000genomes. org/). None of the single SNPs imputed at this locus ($n = 2538$) showed stronger evidence of association with the AD risk than the haplotypes initially identified (Supplementary Figure 1). The 10p13 haplotype region appeared to be fully included
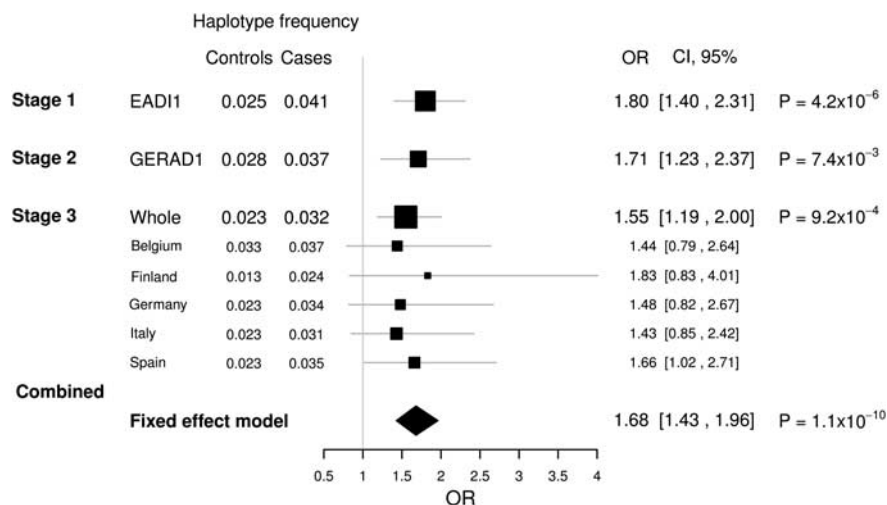


**Figure 1** Haplotypic Odd ratios (ORs) for Alzheimer's disease (AD) risk with the AAC haplotype derived from rs7081208, rs2446581, rs17314229 at Chromosome 10p13 in seven independent European populations.
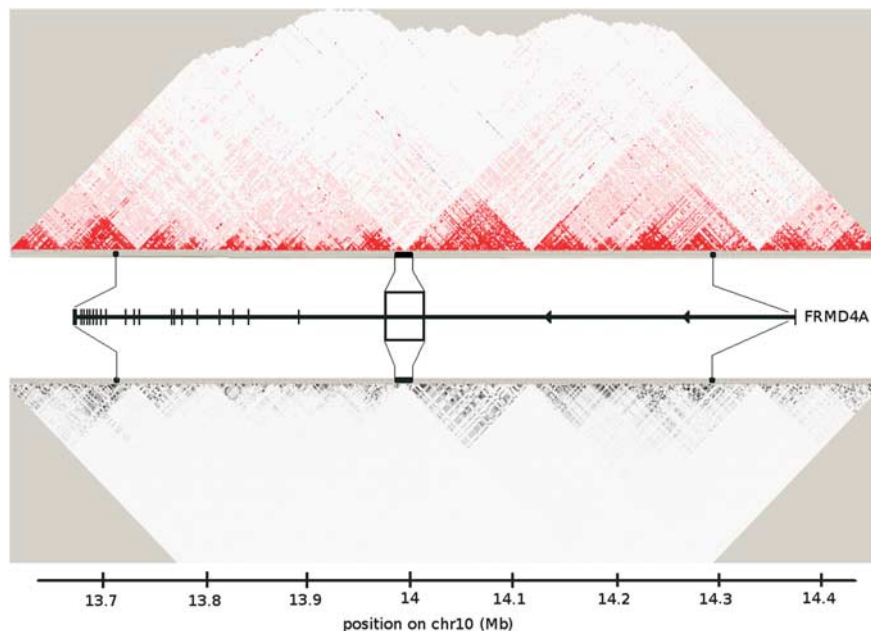
**Figure 2** Linkage disequilibrium map in the *FRMD4A* locus and localization of the region defined by the rs7081208, rs2446581 and rs17314229 at this locus (in bold).

within the FERM domain containing 4A (*FRMD4A*) gene, as indicated by the linkage disequilibrium map of this locus of interest (Figure 2).

We finally explored how *FRMD4A* might be involved in the AD process. According to a recent report, the *FRMD4A* gene was described to interact with Arf6.[24] As this latter protein was reported to control APP processing,[25] we postulated that the *FRMD4A* locus could be associated with endophenotypes susceptible to reflect modulation of the APP metabolism. We accordingly analyzed association of the *FRMD4A* locus with Aβ peptide plasma concentrations in three independent populations of non-demented individuals ($n = 2579$) (Supplementary Table 2) for which *FRMD4A* imputed SNPs, $A\beta_{1-40}$ and $A\beta_{1-42}$ plasma concentrations were available. Strong associations were observed between several *FRMD4A* SNPs and $A\beta_{1-42}/A\beta_{1-40}$ (nine SNPs reaching a significant level after Bonferroni correction, $P = 3.4 \times 10^{-5}$ for 1486 SNPs, Figure 3). The best signal was obtained for rs7921545 (meta-analysed $z$-score β coefficient: 0.12, CI: 95% (0.07–0.017); $P = 5.4 \times 10^{-7}$) and was homogeneous between the three data sets ($P$ for heterogeneity $= 3.7 \times 10^{-1}$, Supplementary Figure 2). Of note, rs2446581 showed nominal association with $A\beta_{1-42}/A\beta_{1-40}$ ($P = 3.1 \times 10^{-2}$, see Supplementary Figure 2). Furthermore, only nominal associations between *FRMD4A* SNPs, $A\beta_{1-40}$ or $A\beta_{1-42}$ were detected (see Supplementary Figure 3).

## Discussion

From a specific haplotype-based GWAS approach, we were able to detect a new genetic susceptibility

factor for AD that could not be identified through usual GWAS analyses. Owing to the high number of association tests performed in this GWHA study (37 330 050, not all of which were independent), a robust replication strategy was necessary. Our three-step approach was thus particularly conservative and we cannot rule out the possibility that we failed to identify other haplotype-based loci associated with AD risk. Nonetheless, our GWHA immediately identified two potential AD susceptibility loci among which the 10p13 locus showed an AAC haplotype that was strongly and consistently associated with AD risk in seven independent European populations. Interestingly, this locus is included in the large AD linkage region regularly identified on chromosome 10.[26]

Additional work will be necessary to determine whether the AD risk-haplotype association indicates an interaction between SNPs or whether they tag non-genotyped, functional variants. However, our GWHA study could have handled several of the inherent limitations of GWAS. As the AAC haplotype is rare (with a mean frequency of 2% in our Caucasian populations), our work suggests the possibility that rare variants may be responsible for the signal detected within the *FRMD4A* gene. This might explain why the locus was not detected (i) in our previous GWA studies based on single-SNPs analyses[5–9] and (ii) through imputation, as SNPs with low frequency and/or SNPs within specific regions with low LD are poorly imputed even when using the 1000 genome data set (Figure 2). Furthermore, we cannot rule out the possibility that the AAC haplotype tags insertion–deletion variants or copy-number
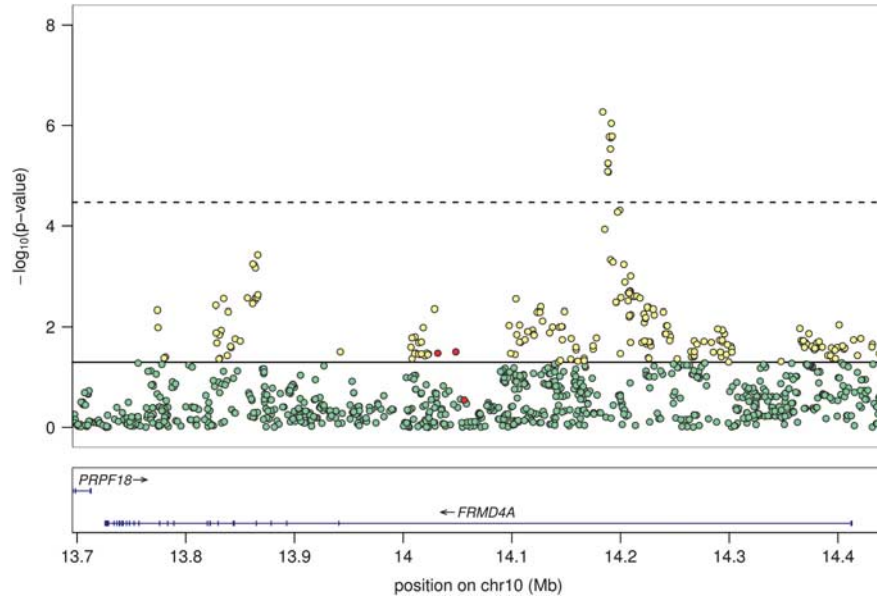
**Figure 3** Association of single-nucleotide polymorphism (SNP) in the *FRMD4A* locus with plasma $A\beta_{1-42}/A\beta_{1-40}$ level following meta-analyses of z-score $\beta$ coefficients under an additive model adjusted for age and gender using three independent healthy populations. SNPs in red are nominally associated with $A\beta$ peptide levels. SNPs in are the three markers defining the AAC haplotype assocated with Alzheimer's disease (AD) risk.

variations—neither of which is captured by imputation. Interestingly, several copy-number variations within the *FRMD4A* locus have been described (http://projects.tcag.ca/variation/).

Little is known about the function of the protein encoded by *FRMD4A* in animal cells. It belongs to the FERM super family, which includes ubiquitous components of the cytocortex involved in cell structure, transport and signalling functions.[27] According to a recent report, the *FRMD4A* gene product may regulate epithelial polarity by interacting with Arf6 and the PAR complex.[24] Interestingly, several other lines of evidence suggest that Arf6 modulates cell polarity in various systems—including neurons. Arf6 reportedly regulates dendritic branching in hippocampal neurons and neurite outgrowth in PC12 cells.[28,29] Finally, Arf6 was recently reported to control APP processing,[27] suggesting that *FRMD4A* could also be implied in this metabolism. This hypothesis is sustained by the observation of an association of the *FRMD4A* locus with plasma $A\beta_{1-42}/A\beta_{1-40,}$ at once reinforcing the plausibility of the association of this gene with AD risk and its potential implication in a subtle control of the APP metabolism. Unfortunately, as the SNPs defining the AAC haplotype were only available by imputation in the Rotterdam and CHS study, search for association of the AAC haplotype with plasma $A\beta_{1-42}/A\beta_{1-40}$ was not possible. Furthermore, it is important to keep in mind that this result is difficult to interpret in terms of AD pathophysiological process. First, it is not known whether plasma $A\beta$ peptides reflect a dynamic equilibrium between the brain, CSF and plasma compartments.[30–34] Second, the source of plasma $A\beta$ species is

not known and the $A\beta$ peptides' physiological functions are still not fully understood.[35,36]

However, taken as a whole, these data suggest that *FRMD4A* could be a relevant candidate gene for AD risk and the basis for another possible pathophysiological pathway for AD.

## Conflict of interest

The authors declare no conflict of interest.

## Acknowledgments

## References

1 Hardy J, Selkoe D. The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics. *Science* 2002; **297**: 353–356.

2 Small SA, Duff K. Linking Abeta and tau in late-onset Alzheimer's disease: a dual pathway hypothesis. *Neuron* 2008; **60**: 534–542.

3 Lambert JC, Amouyel P. Genetic heterogeneity of Alzheimer's disease: complexity and advances. *Psychoneuroendocrinology* 2007; **32**: S62–S70.

4 Genin E, Hannequin D, Wallon D, Sleegers K, Hiltunen M, Combarros O *et al.* APOE and Alzheimer disease: a major gene with semi-dominant inheritance. *Mol Psychiatry* 2011; **16**: 903–907.

5 Lambert JC, Heath S, Even G, Campion D, Sleegers K, Hiltunen M *et al.* Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease. *Nat Genet* 2009; **41**: 1094–1099.

6 Harold D, Abraham R, Hollingworth P, Sims R, Gerrish A, Hamshere ML *et al.* Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. *Nat Genet* 2009; **41**: 1088–1093.

7 Seshadri S, Fitzpatrick AL, Ikram MA, DeStefano AL, Gudnason V, Boada M *et al.* Genome-wide analysis of genetic loci associated with Alzheimer disease. *JAMA* 2010; **303**: 1832–1840.

8 Hollingworth P, Harold D, Sims R, Gerrish A, Lambert JC, Carrasquillo MM *et al.* Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. *Nat Genet* 2011; **43**: 429–435.

9 Naj AC, Jun G, Beecham GW, Wang LS, Vardarajan BN, Buros J *et al.* Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. *Nat Genet* 2011; **43**: 436–441.

10 Lambert JC, Amouyel P. Genetics of Alzheimer's disease: new evidences for an old hypothesis? *Curr Opin Genet Dev* 2011; **21**: 295–301.

11 Trégouët DA, König IR, Erdmann J, Munteanu A, Braund PS, Hall AS *et al.* Genome-wide haplotype association study identifies the SLC22A3-LPAL2-LPA gene cluster as a risk locus for coronary artery disease. *Nat Genet* 2009; **41**: 283–285.

12 McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA work group under the auspices of department of Health and Human services task force on Alzheimer's disease. *Neurology* 1984; **34**: 939–944.

13 Mirra SS, Heyman A, McKeel D, Sumi SM, Crain BJ, Brownlee LM *et al.* The consortium to establish a registry for Alzheiemr's disease (CERAD). Part II. Standardization of the neuropathologic assessment of Alzheimer's disease. *Neurology* 1991; **41**: 479–486.

14 Van Oijen M, Hofman A, Soares HD, Koudstaal PJ, Breteler MM. Plasma Abeta(1-40) and Abeta(1-42) and the risk of dementia: a prospective case-cohort study. *Lancet Neurol* 2006; **5**: 655–660.

15 Lambert JC, Schraen-Maschke S, Richard F, Fievet N, Rouaud O, Berr C *et al.* Association of plasma amyloid beta with risk of dementia: the prospective three-city study. *Neurology* 2009; **73**: 847–853.

16 Lopez OL, Kuller LH, Mehta PD, Becker JT, Gach HM, Sweet RA *et al.* Plasma amyloid levels and the risk of AD in normal subjects in the cardiovascular health study. *Neurology* 2008; **70**: 1664–1671.

17 Mathias RA, Gao P, Goldstein JL, Wilson AF. A graphical assessment of *P*-values from sliding window haplotype tests of association to identify asthma susceptibility loci on chromosome 11q. *BMC Genet* 2006; **7**: 38.

10

18 Yu Z, Schaid DJ. Sequential haplotype scan methods for association analysis. *Genet Epidemiol* 2007; **31**: 553–564.

19 Huang BE, Amos CI, Lin DY. Detecting haplotype effects in genome wide association studies. *Genet Epidemiol* 2007; **31**: 803–812.

20 Tregouet DA, Ricard S, Nicaud V, Arnould I, Soubigou S, Rosier M *et al.* In-depth haplotype analysis of ABCA1 gene polymorphisms in relation to plasma ApoA1 levels and myocardial infarction. *Arterioscler Thromb Vasc Biol* 2004; **24**: 775–781.

21 Tregouet DA, Escolano S, Tiret L, Mallet A, Golmard JL. A new algorithm for haplotype-based association analysis: the stochastic-EM algorithm. *Ann Hum Genet* 2004; **68**: 165–177.

22 Tregouet DA, Garelle V. A new JAVA interface implementation of THESIAS: testing haplotype effects in association studies. *Bioinformatics* 2007; **23**: 1038–1039.

23 Gagliardi F, Jones B, Grey F, Bégin ME, Heikkurinen M. Building an infrastructure for scientific grid computing: status and goals of the EGEE project. *Philos Transact A Math Phys Eng Sci* 2005; **363**: 1729–1742.

24 Ikenouchi J, Umeda M. FRMD4A regulates epithelial polarity by connecting Arf6 activation with the PAR complex. *Proc Natl Acad Sci USA* 2010; **107**: 748–753.

25 Sannerud R, Declerck I, Peric A, Raemaekers T, Menendez G, Zhou L *et al.* ADP ribosylation factor 6 (ARF6) controls amyloid precursor protein (APP) processing by mediating the endosomal sorting of BACE1. *Proc Natl Acad Sci USA* 2011; **108**: E559–E568.

26 Hamshere ML, Holmans PA, Avramopoulos D, Bassett SS, Blacker D, Bertram L *et al.* Genome-wide linkage analysis of 723 affected relative pairs with late-onset Alzheimer's disease. *Hum Mol Genet* 2007; **16**: 2703–2712.

27 Tepass U. FERM proteins in animal morphogenesis. *Curr Opin Genet Dev* 2009; **19**: 357–367.

28 Hernández-Deviez DJ, Casanova JE, Wilson JM. Regulation of dendritic development by the ARF exchange factor ARNO. *Nat Neurosci* 2002; **5**: 623–624.

29 Albertinazzi C, Za L, Paris S, de Curtis I. ADP-ribosylation factor 6 and a functional PIX/p95-APP1 complex are required for Rac1B-mediated neurite outgrowth. *Mol Biol Cell* 2003; **14**: 1295–1307.

30 Mehta PD, Pirttila T, Patrick BA, Barshatzky M, Mehta SP. Amyloid beta protein 1-40 and 1-42 levels in matched cerebrospinal fluid and plasma from patients with Alzheimer disease. *Neurosci Lett* 2001; **304**: 102–106.

31 Vanderstichele H, Van Kerschaver E, Hesse C, Davidsson P, Buyse MA, Andreasen N *et al.* Standardization of measurement of beta-amyloid(1-42) in cerebrospinal fluid and plasma. *Amyloid* 2000; **7**: 245–258.

32 Kawarabayashi T, Younkin LH, Saido TC, Shoji M, Ashe KH, Younkin SG. Age-dependent changes in brain, CSF, and plasma amyloid (beta) protein in the Tg2576 transgenic mouse model of Alzheimer's disease. *J Neurosci* 2001; **21**: 372–381.

33 Ghersi-Egea JF, Gorevic PD, Ghiso J, Frangione B, Patlak CS, Fenstermacher JD. Fate of cerebrospinal fluid-borne amyloid beta-peptide: rapid clearance into blood and appreciable accumulation by cerebral arteries. *J Neurochem* 1996; **67**: 880–883.

34 Giedraitis V, Sundelöf J, Irizarry MC, Gårevik N, Hyman BT, Wahlund LO *et al.* The normal equilibrium between CSF and plasma amyloid beta levels is disrupted in Alzheimer's disease. *Neurosci Lett* 2007; **427**: 127–131.

35 Soscia SJ, Kirby JE, Washicosky KJ, Tucker SM, Ingelsson M, Hyman B *et al.* The Alzheimer's disease-associated amyloid beta-protein is an antimicrobial peptide. *PLoS One* 2010; **5**: e9505.

36 Lambert JC, Dallongeville J, Ellis KA, Schraen-Maschke S, Lui J, Laws S *et al.* Association of plasma Aβ peptides with blood pressure in the elderly. *PLoS One* 2011; **6**: e18536.

Supplementary Information accompanies the paper on the Molecular Psychiatry website (http://www.nature.com/mp)